



УНИВЕРЗИТЕТ У НОВОМ САДУ
ПРИРОДНО-МАТЕМАТИЧКИ
ФАКУЛТЕТ
ДЕПАРТМАН ЗА МАТЕМАТИКУ
И ИНФОРМАТИКУ



Филип Васић, 50/19

Домаћи задатак из предмета Социјалне мреже
за школску 2021/2022 годину

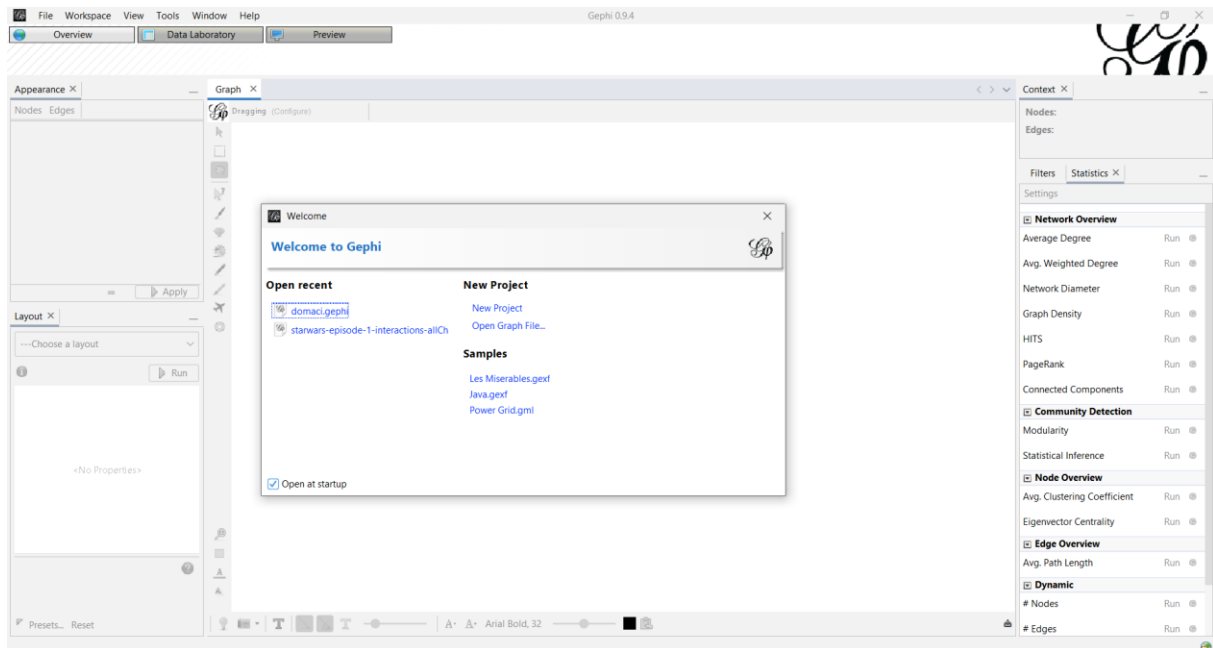
**„Изјављујем под пуном одговорношћу да сам задатке
решавао самостално”**

Нови Сад, 2022.

Gephi

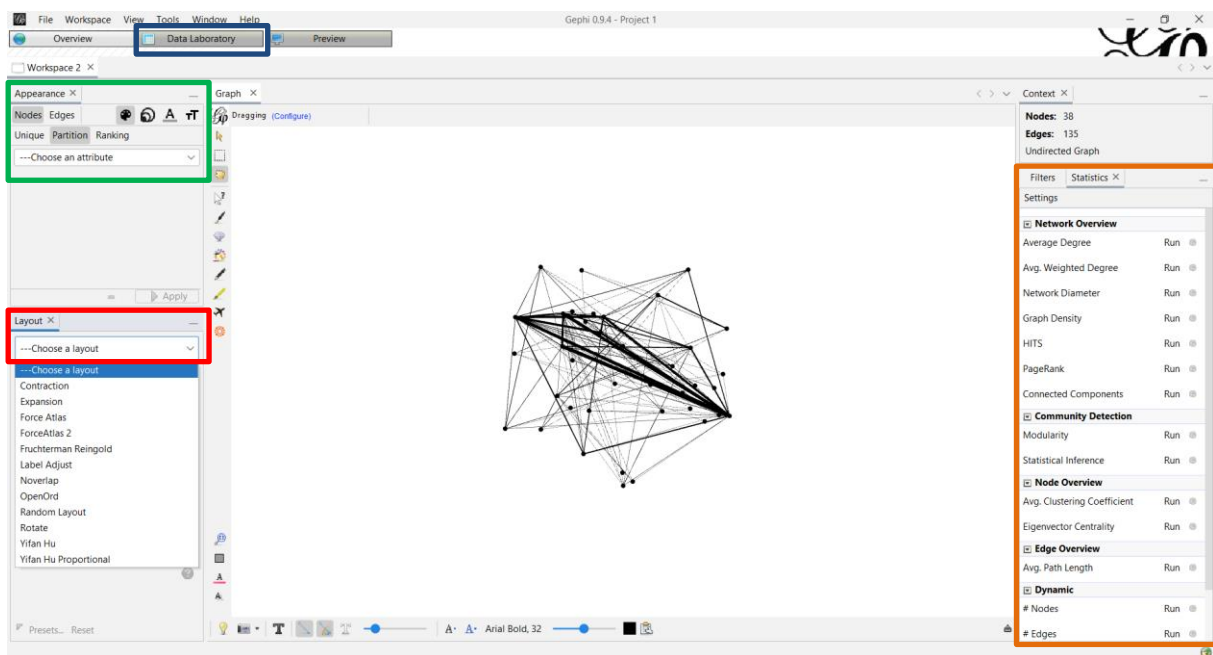
Gephi је алат за анализу и визуализацију великих графова. Овај алат ће у наставку бити коришћен за анализу графа који описује интеракције јунака Ратова Звезда у епизоди 1.

Помоћу алата одрађена је анализа неких основних метрика као што су просечни степен чворова, густина мреже, дијаметар мреже, просечна дистанца удаљености између чворова и просечан коефицијент кластерисања.



Слика 1

На слици 1 је приказан изглед алата. На почетку имамо опцију да креирамо нови пројекат или да отворимо неки на којем смо већ радили. Након што изаберемо шта желимо приказати се изглед графа. Пре учитавања графа потребно је одабрати да ли је граф усмерен или неусмерен.



Слика 2

На слици 2 приказан је изглед окружења након избора графа за анализу. Примећујемо да приказани граф није баш прегледан и да не можемо много тога видети на оваквом приказу.

Са леве стране налази се одељак *layout*, који је на слици уоквирен црвеном бојом, у њему можемо подешавати неке опције како би граф био боље организован и прегледан. На пример имамо опције *expansion* и *contraction* које служе за зумирање односно умањивање приказа. Такође постоје опције *force atlas* и *force atlas 2* које примају низ параметара и дају поприлично леп приказ графа.

Такође са леве стране постоји одељак *appearance*, који је на слици уоквирен зеленом бојом, и у њему можемо партиционисати или рангирати чворове односно везе по неком критеријуму. На пример можемо рангирати чворове по величини тако да већи чворови имају већи степен а мањи чворови имају мањи степен.

Са десне стране се налазе одељци *filters* и *statistics* уоквирени наранџастом бојом. У одељку *filters* можемо филтрирати граф, на пример можемо одредити k-core, док у одељку *statistics* можемо одредити неке основне метрике за наш граф. То радимо тако што поред метрике коју желимо кликнемо на *run*, затим се добија извештај у којем се виде резултати метрике, такође се у неким метрикама добијају и прикази кроз графике.

На врху странице, на слици уоквирено плавом бојом налази се *data laboratory*, ту можемо видети све статистике које су урађене приказане у једној табели, где можемо сортирати чворове по жељи. Како изгледа *data laboratory* приказано је на слици 3.

FileWorkspaceViewToolsWindowHelp

OverviewData LaboratoryPreview

Workspace 1

Data Table

NodesEdgesConfigurationAdd nodeAdd edgeSearch/ReplaceImport SpreadsheetExport tableMore actions

Filter:Id

Id	Label	Interval	name	value	colour	Degree	Eigenvector Centrality	Authority	Hub	Modularity Class	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality	Component ID
0	0	R2-D2	33	#bde0f6	6	0.358767	0.142934	0.142935	0	3.0	0.486842	0.545045	0.0	0	
1	1	QUI-GON	61	#4f4fb1	26	1.0	0.394105	0.394104	0	2.0	0.770833	0.851351	0.322614	0	
2	2	NUTE GUNR...	19	#808080	10	0.311073	0.11637	0.11637	1	3.0	0.552239	0.621622	0.125431	0	
3	3	PK-4	3	#808080	1	0.018829	0.006721	0.006721	1	4.0	0.336364	0.358108	0.0	0	
4	4	TC-14	5	#808080	5	0.190586	0.07265	0.07265	1	3.0	0.5	0.545045	0.073106	0	
5	5	OB1-WAN	34	#48d1cc	13	0.629926	0.249234	0.249234	3	3.0	0.569231	0.657658	0.044071	0	
6	6	DOFINE	4	#808080	3	0.055694	0.0189	0.0189	1	4.0	0.373737	0.421171	0.001802	0	
7	7	RUNE	11	#808080	4	0.087071	0.029986	0.029986	1	4.0	0.389474	0.445946	0.003554	0	
8	8	TEY HOW	5	#808080	3	0.047402	0.015288	0.015288	1	4.0	0.366337	0.412162	0.000751	0	
9	9	EMPEROR	14	#191970	11	0.415138	0.159028	0.159028	1	3.0	0.560606	0.635135	0.100236	0	
10	10	CAPTAIN PA...	20	#808080	9	0.526789	0.208122	0.208123	2	3.0	0.544118	0.608108	0.009238	0	
11	11	SIO BIBBLE	8	#808080	7	0.412196	0.161264	0.161264	2	3.0	0.528571	0.581081	0.00428	0	
12	12	JAR JAR	36	#9a9a00	18	0.758114	0.299387	0.299386	2	3.0	0.627119	0.72973	0.145836	0	
13	13	TARPAIS	4	#808080	1	0.069864	0.027696	0.027696	2	4.0	0.389474	0.421171	0.0	0	
14	14	BOSS NASS	5	#808080	6	0.423702	0.168422	0.168423	2	3.0	0.5	0.554054	0.0	0	
15	15	PADME	31	#d1aadd	18	0.801401	0.316987	0.316986	2	3.0	0.649123	0.738739	0.103421	0	
16	16	RIC OUE	12	#808080	7	0.375808	0.148554	0.148554	3	3.0	0.506849	0.567568	0.023992	0	
17	17	WATTO	8	#808080	6	0.358767	0.142934	0.142935	0	3.0	0.486842	0.545045	0.0	0	
18	18	ANAKIN	41	#ce3b59	23	0.888764	0.352781	0.352781	3	3.0	0.649123	0.783784	0.169884	0	
19	19	SEBULBA	4	#808080	6	0.400979	0.159712	0.159712	2	3.0	0.5	0.554054	0.0	0	
20	20	JIRA	4	#808080	3	0.247793	0.098417	0.098418	0	3.0	0.468154	0.504505	0.0	0	
21	21	SHMI	11	#808080	8	0.495258	0.197357	0.197358	2	3.0	0.513889	0.581081	0.001502	0	
22	22	C-3PO	6	#ff0700	6	0.358767	0.142934	0.142935	0	3.0	0.486842	0.545045	0.0	0	
23	23	DARTH MA...	6	#808080	4	0.095481	0.033456	0.033456	1	4.0	0.389474	0.445946	0.000501	0	
24	24	KITSTER	5	#808080	9	0.490439	0.195352	0.195353	0	3.0	0.521127	0.594595	0.006194	0	
25	25	WALD	4	#808080	5	0.308161	0.122299	0.1223	0	3.0	0.493333	0.540541	0.002016	0	
26	26	FODE/BEED	12	#808080	3	0.181586	0.072361	0.072362	2	4.0	0.430233	0.475225	0.0	0	
27	27	JABBA	4	#808080	7	0.416555	0.165839	0.16584	2	3.0	0.506849	0.567568	0.005309	0	
28	28	GREEDO	3	#808080	3	0.202846	0.080407	0.080407	0	3.0	0.4625	0.5	0.0	0	

Add columnMerge columnsDelete columnClear columnCopy data to other columnFill column with a valueDuplicate columnCreate a boolean column from regex matchCreate column with list of regex matching groupsNegate boolean valuesConvert column to dynamic

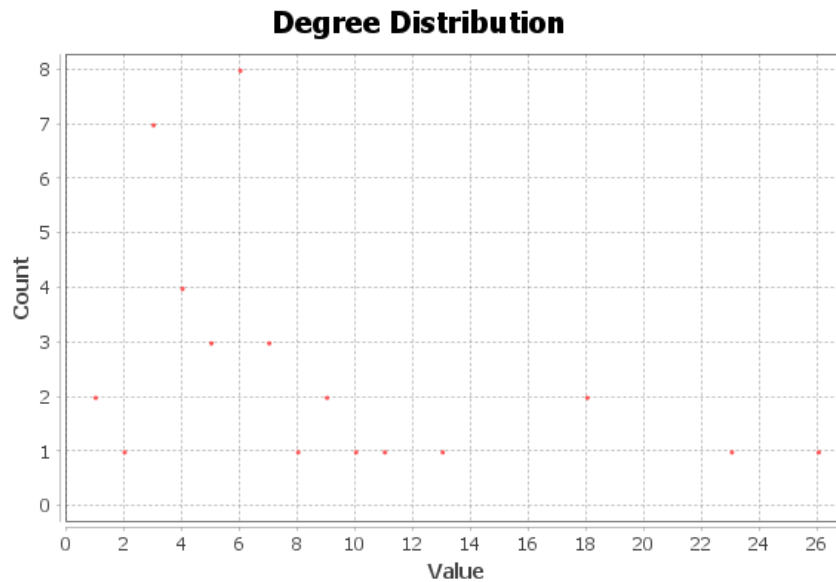
Слика 3

У наставку ће на примеру бити приказано више о свему овоме.

Основне метрике

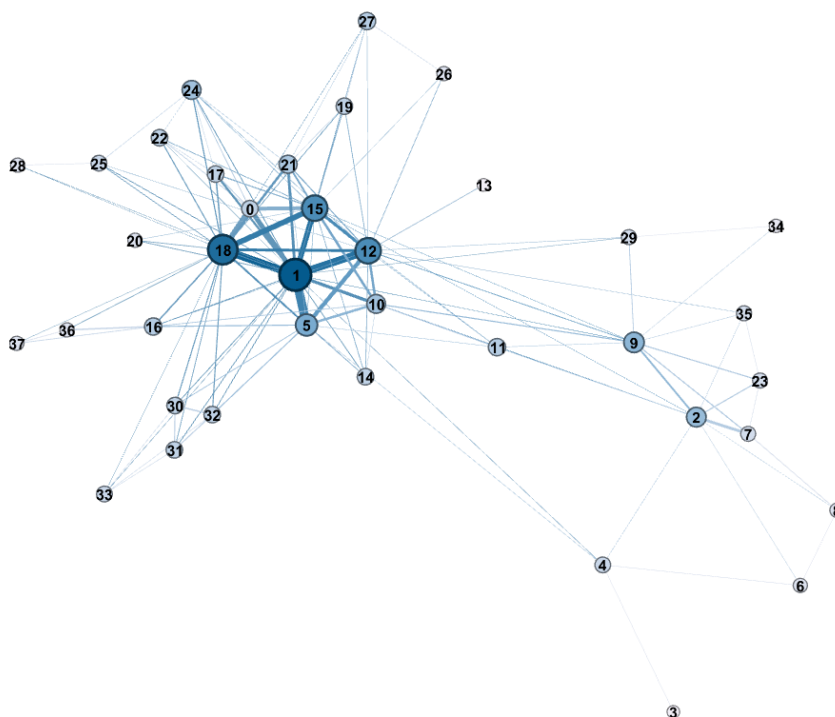
Граф има 38 чворова и 135 веза

Просечан степен чворова: 7,105



Слика 4

На слици 4 приказан је график који представља извештај метрике. На графику се види да највећи број чворова има степен у интервалу [3, 7], тј. највише 8 чворова има степен 6 док 7 чворова има степен 3. У мрежи се налазе два висећа чвора, са степеном 1, нема изолованих чворова, са степеном 0, а највећи степен је 26.

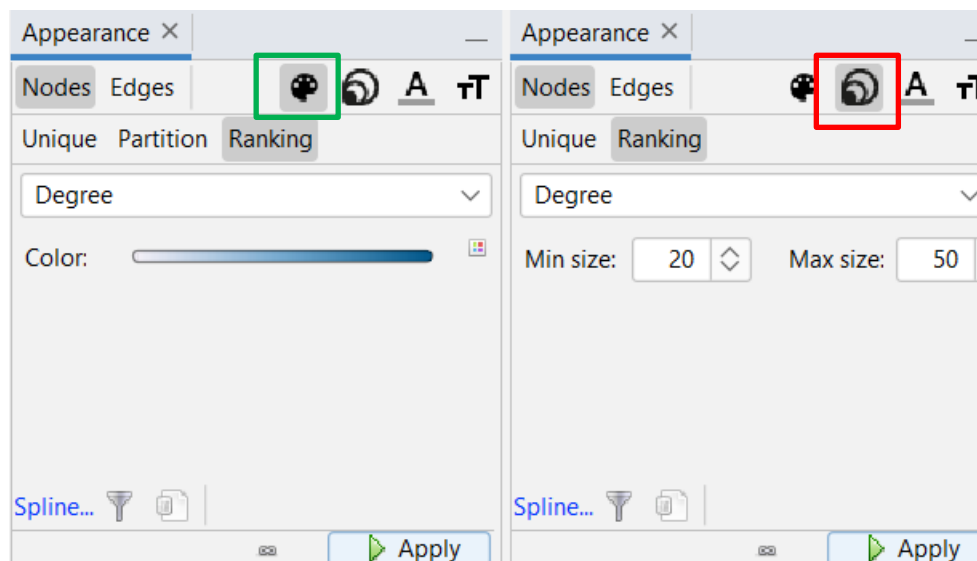


Слика 5

На слици 5 су чворови са већим степеном приказани тамнијом нијансом плаве и већи су.

Овакав приказ се добија тако што у одељку *appearance* кликнемо на *nodes* а затим на *color*, на слици 6 уоквирено зеленом бојом, затим кликнемо на *ranking*, у менију изаберемо *degree* и одговарајући спектар боја. Кликнемо на *apply*.

Величину подешавамо на сличан начин, уместо *color* бирамо опцију *size*, на слици уоквирено црвеном бојом, поново у менију изаберемо *degree* и подесимо минималну и максималну величину чвора, у овом случају су то вредности 20 и 50.

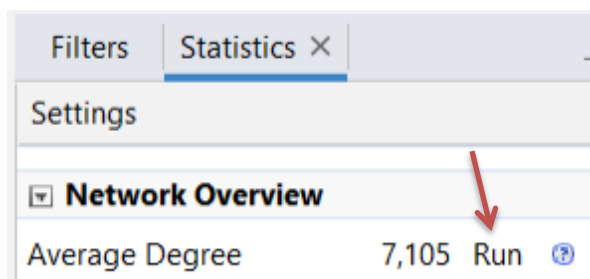


Слика 6

Чвор са највећим степеном је чвор 1.

Мрежа је ретка. Укупан број чворова је 38, мрежа је ретка јер је $7,105 \ll 37$

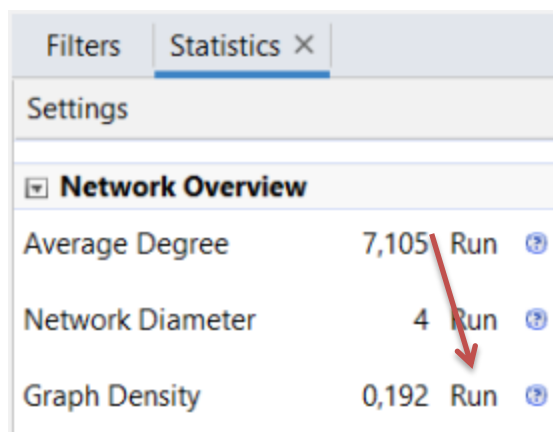
Резултати за ову метрику су добијени тако што се у одељку *statistics* кликне на дугме *run* поред метрике *average degree* што је приказано на слици 7.



Слика 7

Густина мреже: 0,192

Резултати за ову метрику су добијени тако што се у одељку *statistics* кликне на дугме *run* поред метрике *graph density* што је приказано на слици 8.

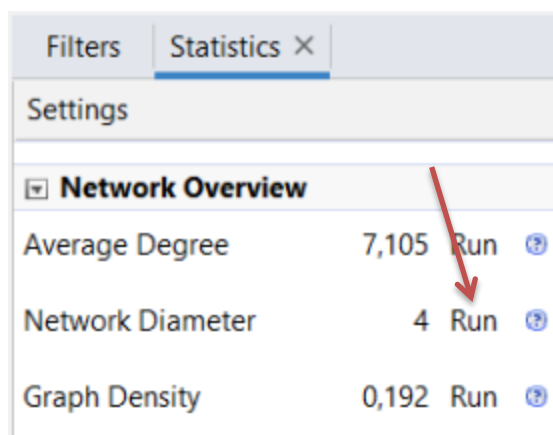


Слика 8

Дијаметар мреже: 4

Резултати за ову метрику су добијени тако што се у одељку *statistics* кликне на дугме *run* поред метрике *network diameter* што је приказано на слици 9.

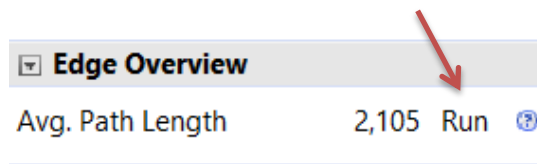
Код ове метрике ћемо након клика на *run* добити нови прозор где можемо да изаберемо да ли је граф усмерен или неусмерен што ће већ бити изабрано ако смо на почетку при отварању пројекта изабрали неку од те две опције. Такође можемо чекирати поље *normalize centralities in [0, 1]* што се односи на централности које ће такође бити израчунате у склопу ове метрике, више о томе у наредном делу задатка.



Слика 9

Просечна дистанца између чворова: 2,105

Резултати за ову метрику су добијени тако што се у одељку *statistics* кликне на дугме *run* поред метрике *average path length* што је приказано на слици 10. Такође резултат ове метрике се добија и у претходном кораку преко дијаметра, тако да ће ова метрика већ бити израчуната.

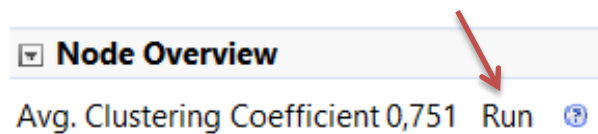


Слика 10

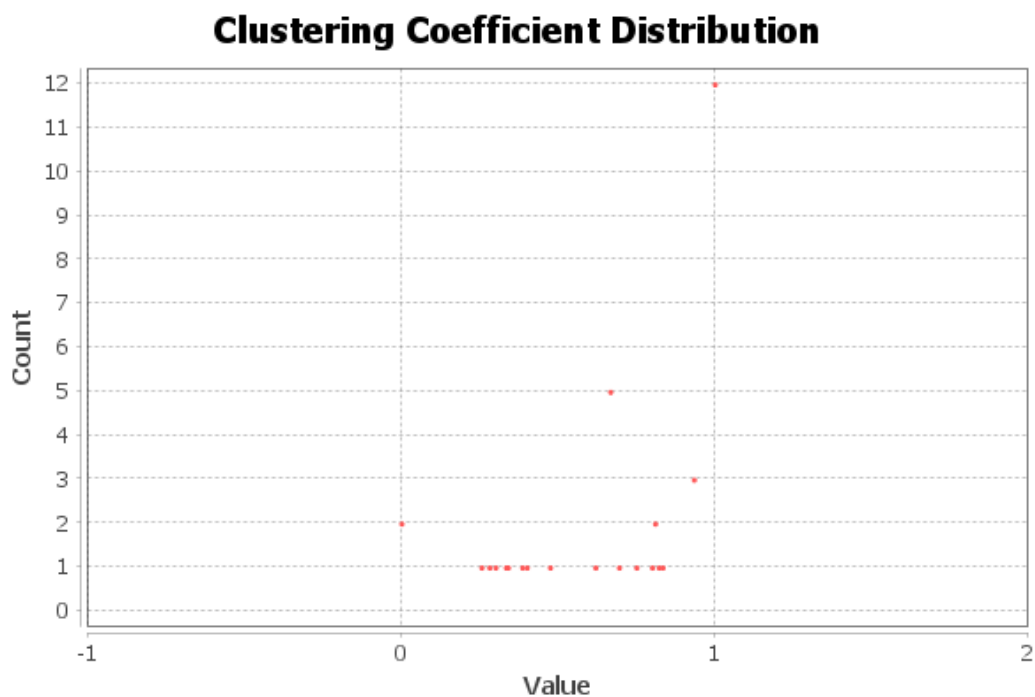
Просечан коефицијент кластерисања: 0,751

Резултати за ову метрику су добијени тако што се у одељку *statistics* кликне на дугме *run* поред метрике *average clustering coefficient* што је приказано на слици 11.

Коефицијент кластерисања представља густину ево мреже без ево чвора што је еквивалентно вероватноћи да су два суседа чвора суседи међусобно.



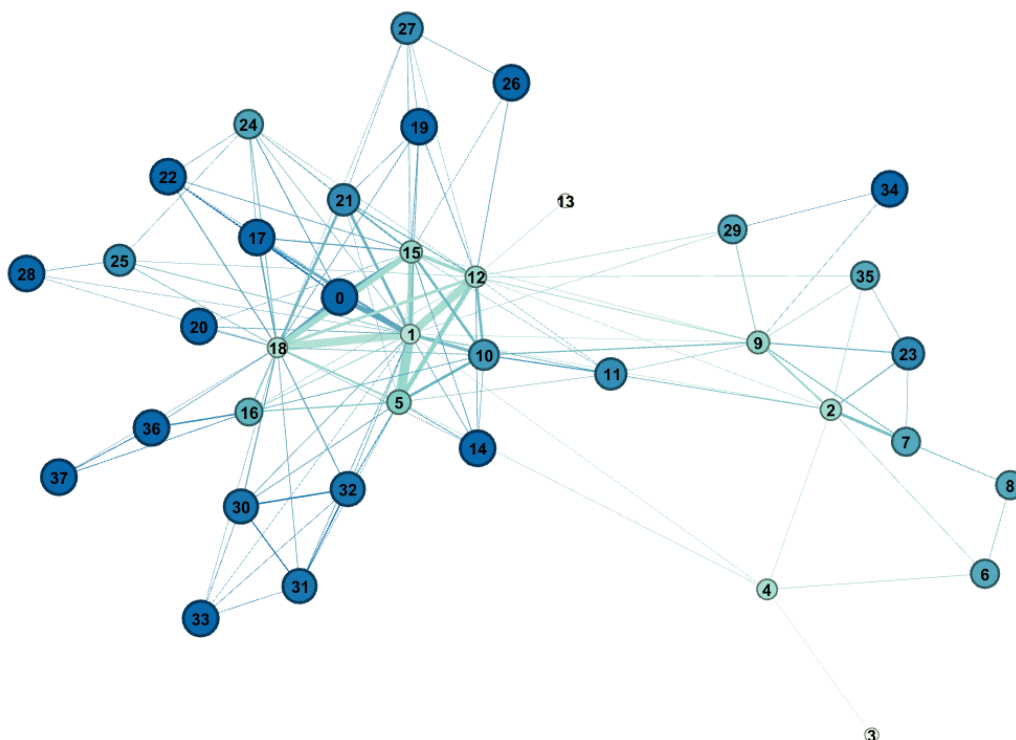
Слика 11



Слика 12

На слици 12 приказан је график метрике за рачунање просечног коефицијента кластерисања. На графику се види коефицијент кластерисања за сваки чвор посебно.

Број троуглова је 212.



Слика 13

На слици 12 су чворови са већим коефицијентом кластерисања приказани тамнијом нијансом плаве и већи су од оних са мањим коефицијентом.

Оваква приказ се добија на исти начин, као што је већ описано код метрике за просечну вредност степена чвора, разлика се само што се у менију бира опција *clustering coefficient*.

Рангирање чворова по метрикама централности

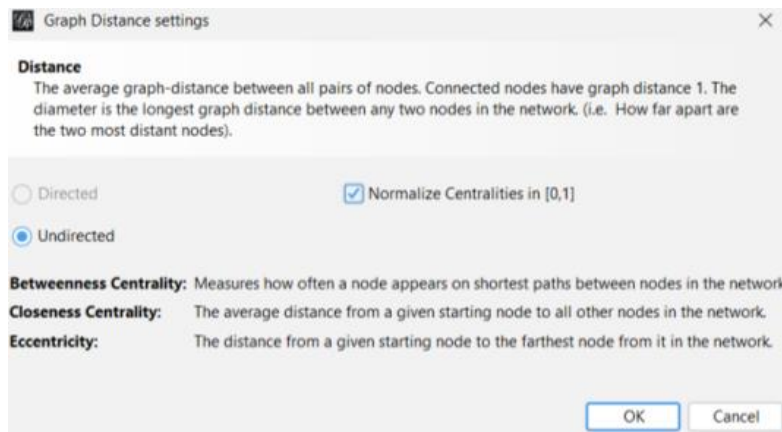
Id	Label	name	value	Degree ▾
1	1	QUI-GON	61	26
18	18	ANAKIN	41	23
12	12	JAR JAR	36	18
15	15	PADME	31	18
5	5	OBI-WAN	34	13
9	9	EMPEROR	14	11
2	2	NUTE GUNR...	19	10
10	10	CAPTAIN PA...	20	9
24	24	KITSTER	5	9
21	21	SHMI	11	8

Слика 14

На слици 14 је приказано 10 чворова са нејвећим степеном.

Betweenness centrality

Резултат за ову метрику као и за *closeness centrality* добија се тако што покренемо рачунање метрике за дијаметар у графу, што смо видели како се ради у претходном делу задатка. Када то урадимо отвара се прозор приказан на слици 15. Чекирамо опцију *normalize centralities in [0,1]* да вредности буду централизоване у том интервалу и кликнемо на *OK*.



Слика 15

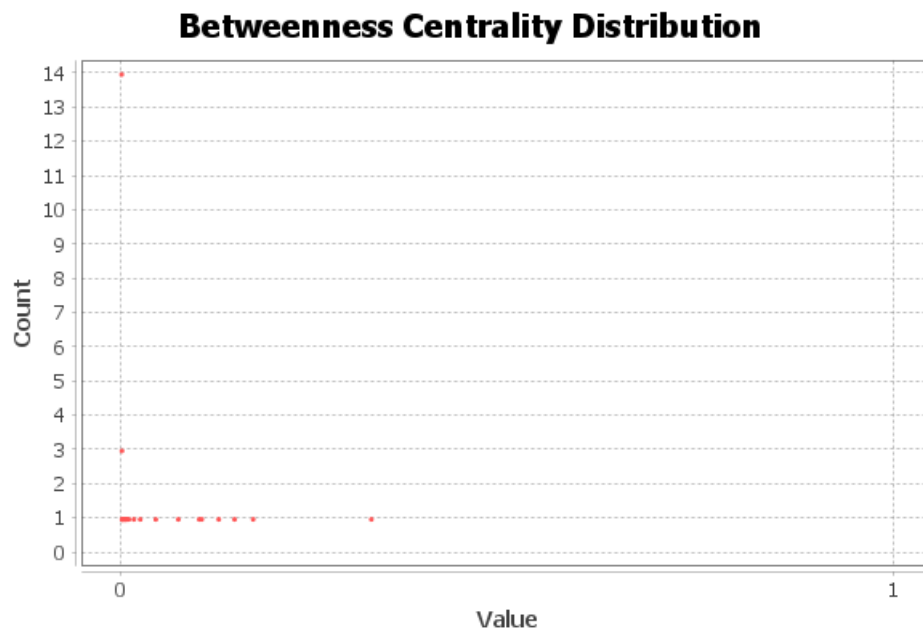
Id	Label	name	value	Degree	Betweenness Centra... ▾
1	1	QUI-GON	61	26	0.322614
18	18	ANAKIN	41	23	0.169884
12	12	JAR JAR	36	18	0.145836
2	2	NUTE GUNR...	19	10	0.125431
15	15	PADME	31	18	0.103421
9	9	EMPEROR	14	11	0.100236
4	4	TC-14	5	5	0.073186
5	5	OBI-WAN	34	13	0.044071
16	16	RIC OLIE	12	7	0.023992
29	29	VALORUM	4	4	0.015785

Слика 16

На слици 16 приказано је 10 чворова који имају највећу највећом *betweenness centrality* и њихови степени. На слици се види да чворови са највећим степеном имају највеће вредности и за ову метрику.

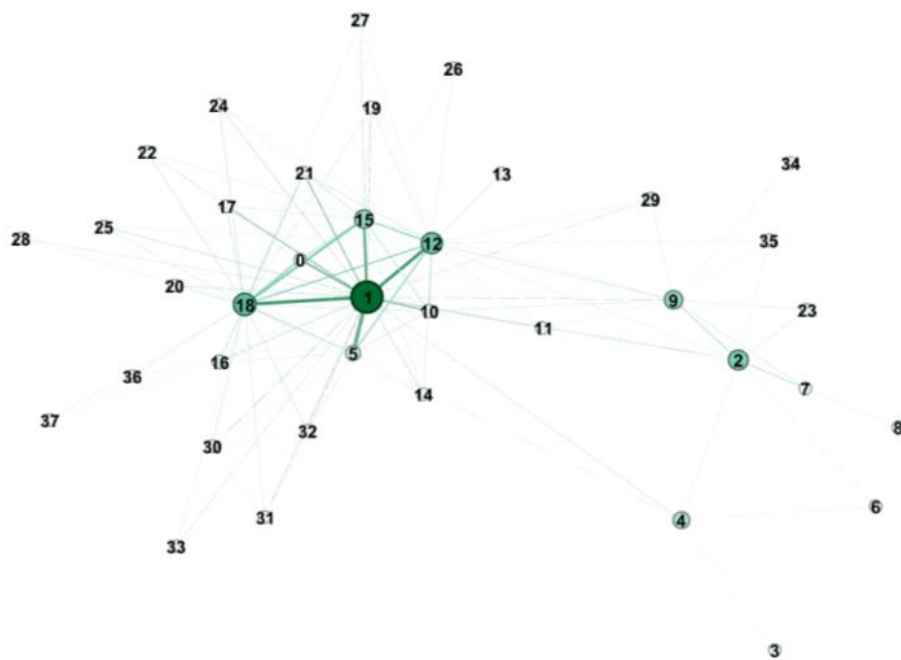
Међутим видимо да чвор са ИД бројем 4 који има степен 5, има већу вредност од чвора са ИД бројем 5 који има степен 13. То значи да чвор број 5 има више веза са другим чворовима али су те везе мање битне од веза чвора број 4, тј. мање битне за проток информација.

На слици 17 се јасније види распоред чворова.



Слика 17

На слици 17 приказан је график метрике. На њему се види да највећи број чворова има вредности блиске 0.



Слика 18

На слици 18 дат је графички приказ графа и веза у њему с тим да су чворови са већом *betweenness centrality* приказани тамнијом нијансом зелене боје и већи су.

Овакав приказ се добија на исти начин као и у претходним случајевима, разлика је поново само у избору опције у менију, овде бирамо *betweenness centrality*.

Closeness centrality

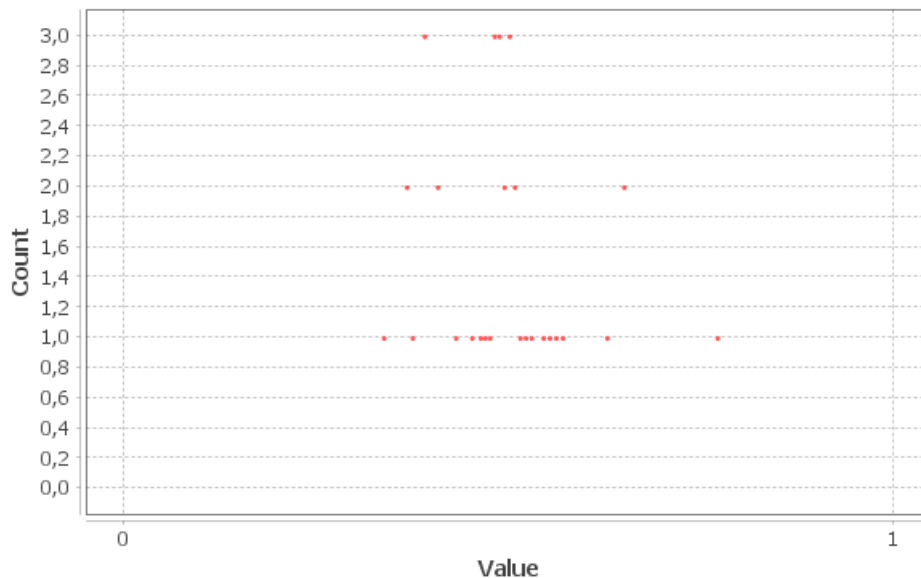
Id	Label	name	value	Degree	Closeness Centra...
1	1	QUI-GON	61	26	0.770833
18	18	ANAKIN	41	23	0.649123
15	15	PADME	31	18	0.649123
12	12	JAR JAR	36	18	0.627119
5	5	OBI-WAN	34	13	0.569231
9	9	EMPEROR	14	11	0.560606
2	2	NUTE GUNR...	19	10	0.552239
10	10	CAPTAIN PA...	20	9	0.544118
11	11	SIO BIBBLE	8	7	0.528571
24	24	KITSTER	5	9	0.521127

Слика 19

На слици 19 приказано је 10 чворова који имају највећу *closeness centrality* и њихови степени. Највећу вредност као и код *betweenes centrality* имају чворови који имају највеће степене с тим да је код ове метрике тај однос са степеном још више изражен што је и логично у једној оваквој мрежи пошто ова метрика даје веће вредности чворовима који се налазе у близини више других чворова.

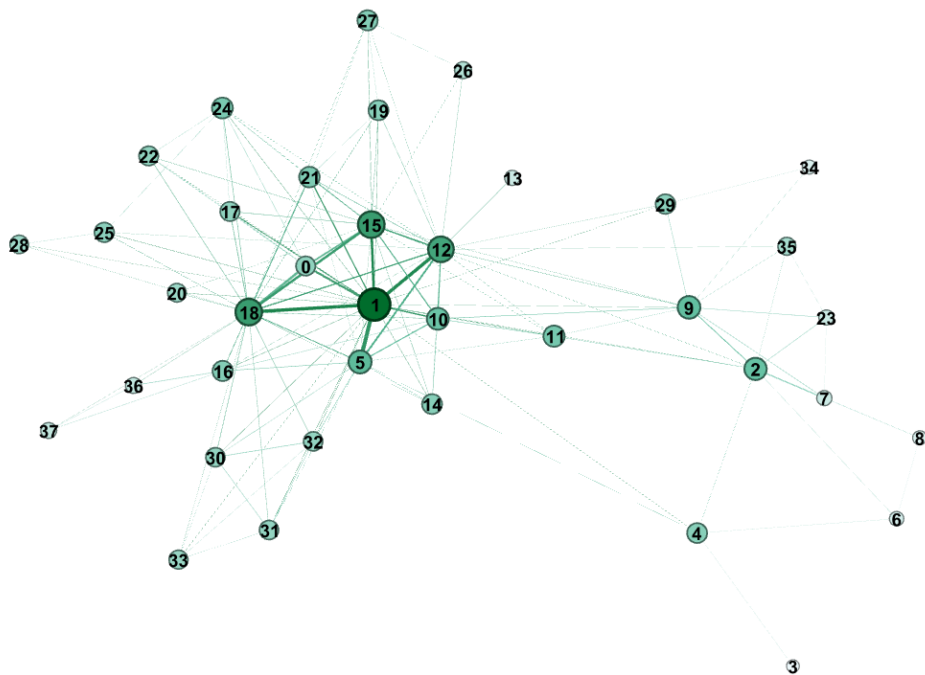
У првих 10 једини изузетак је чвор 11 који има степен 7 а за мало има већу вредност метрике од чвора 24 са степеном 9, то значи да чвор 11 повезује различите кластере унутар графа, што се може видети на слици 11.

Closeness Centrality Distribution



Слика 20

На слици 20 приказан је график метрике. На њему се види да највећи број чворова има вредности у интервалу [0.4, 0.6]



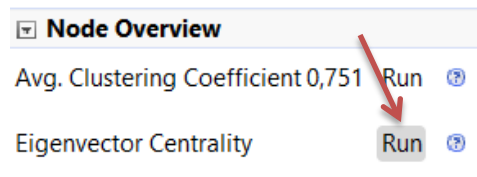
Слика 21

На слици 21 дат је графички приказ графа и веза у њему с тим да су чворови са већом *closeness centrality* приказани тамнијом нијансом зелене боје и већи су.

Овакав приказ се добија на исти начин као и у претходним случајевима, разлика је поново само у избору опције у менију, овде бирамо *closeness centrality*.

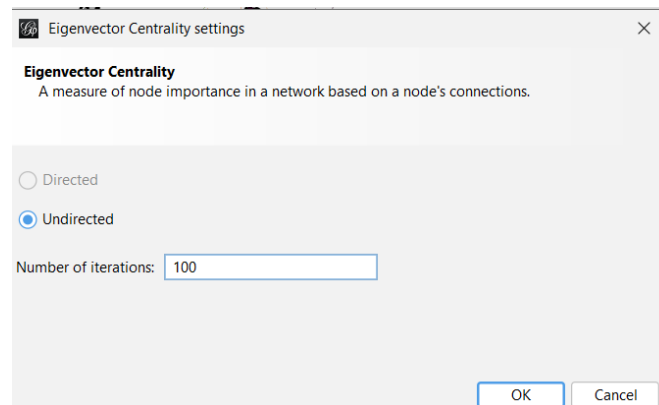
Eigenvector centrality

Eigenvector centrality се добија тако што у одељку *statistics* кликнемо на *run* поред *eigenvector centrality* што је приказано на слици 22.



Слика 22

Након клика на *run* отвара се прозор где можемо да бирамо да ли је граф усмерен или неусмерен али као и код претходних метрика већ је изабрано да је неусмерен јер је тако задато при отаварању фајла. Такође добија се поље у које се може уписати број итерација, default вредност је 100, ми можемо уписати и већу и мању. Како изгледа прозор приказано је на слици 23. Након што је све изабрано кликне се на дугме ОК.

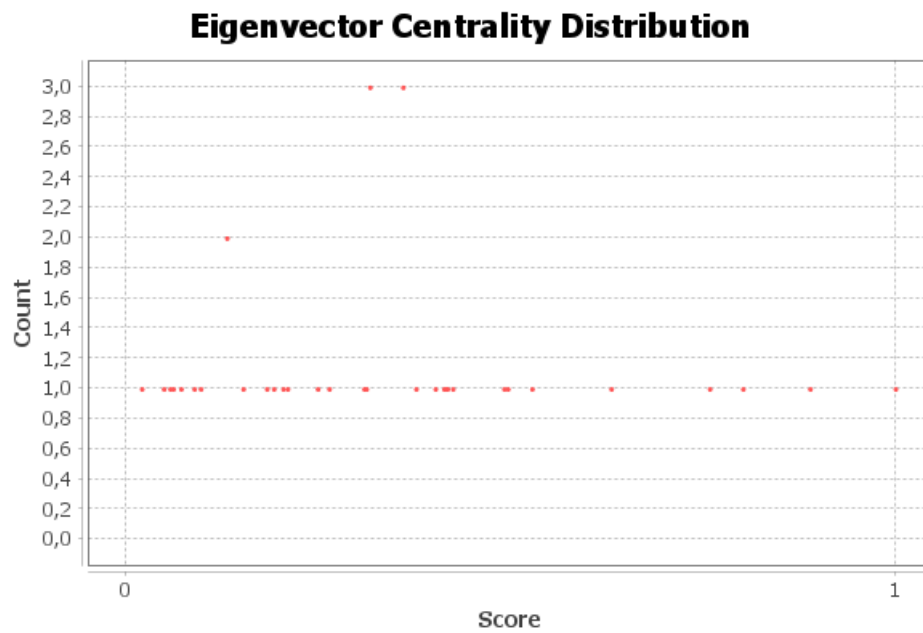


Слика 23

На слици 24 приказано је 10 чворова који имају највећу *eigenvector centrality* на 100 итерација и њихови степени. Као и код *betweeness* и *closeness centrality* највеће вредности за ову метрику имају чворови са највећим степеном.

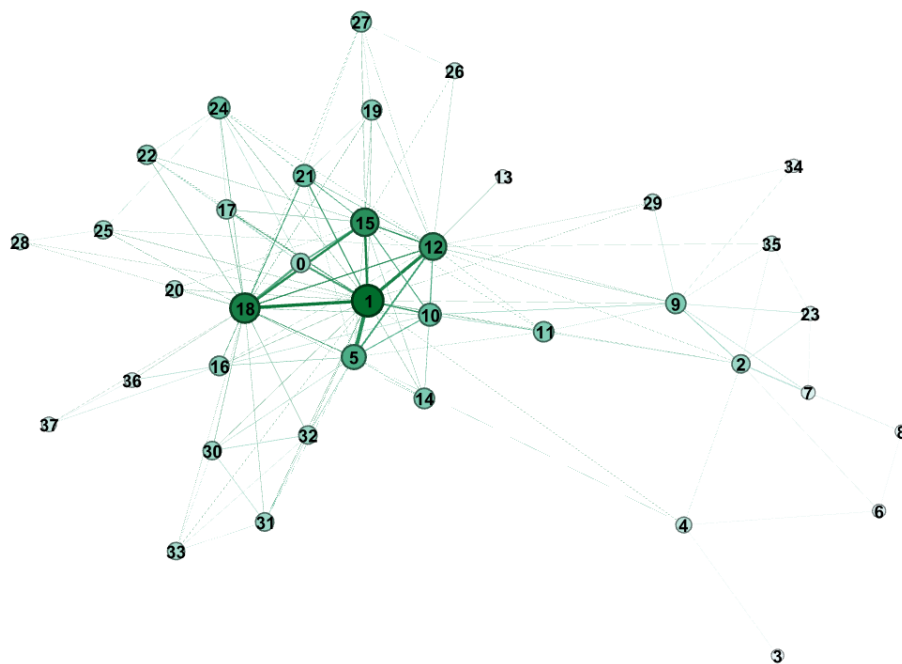
Id	Label	name	value	Degree	Eigenvector Centra...
1	1	QUI-GON	61	26	1.0
18	18	ANAKIN	41	23	0.888764
15	15	PADME	31	18	0.801401
12	12	JAR JAR	36	18	0.758114
5	5	OBI-WAN	34	13	0.629926
10	10	CAPTAIN PA...	20	9	0.526789
21	21	SHMI	11	8	0.495258
24	24	KITSTER	5	9	0.490439
14	14	BOSS NASS	5	6	0.423702
27	27	JABBA	4	7	0.416555

Слика 24



Слика 25

На слици 25 приказан је график метрике.



Слика 26

На слици 26 дат је графички приказ графа и веза у њему с тим да су чворови са већом eigenvector centrality приказани тамнијом нијансом зелене боје и већи су.

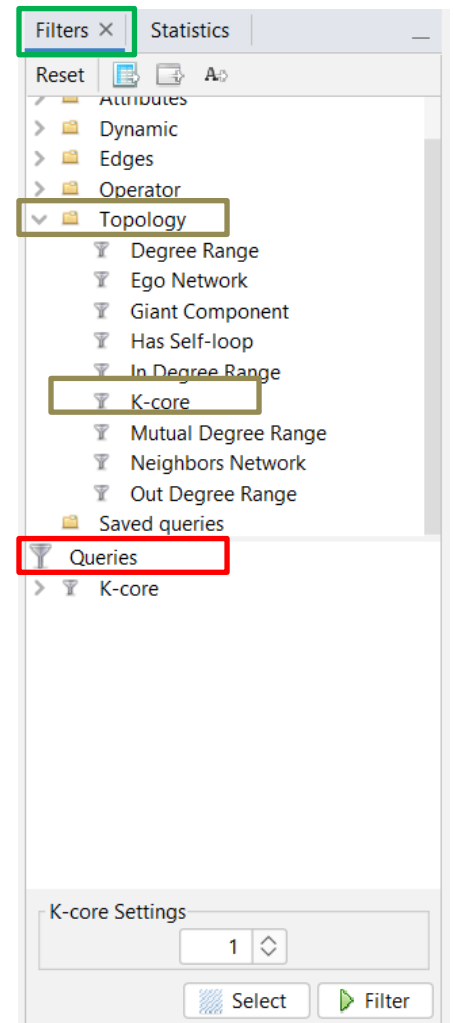
Овакав приказ се добија на исти начин као и у претходним случајевима, разлика је поново само у избору опције у менију, овде бирамо *eigenvector centrality*.

K-core

У десном делу прозора поред одељка *statistics* налази се и одељак *filters*, на слици 27 уоквирен зеленом бојом. Ту можемо бирати филтере по којима ћемо филтрирати чворове у графу.

K-core филтер са налази у одељку *topology*. Да применимо филтер потребно је селектовати га и превући га у одељак *queries* који је на слици уоквирен црвеном бојом. Након тога кликом на дугме филтер можемо филтрирати чворове тако да само они буду приказани и само се они виде у одељку *data laboratory*, а такође можемо их и само селектовати кликом на дугме *select*.

У пољу *k-core settings* можемо подешавати шел индекс тако добијати *2-core*, *3-core* и слично. Такође тако можемо добити и максималан шел индекс.



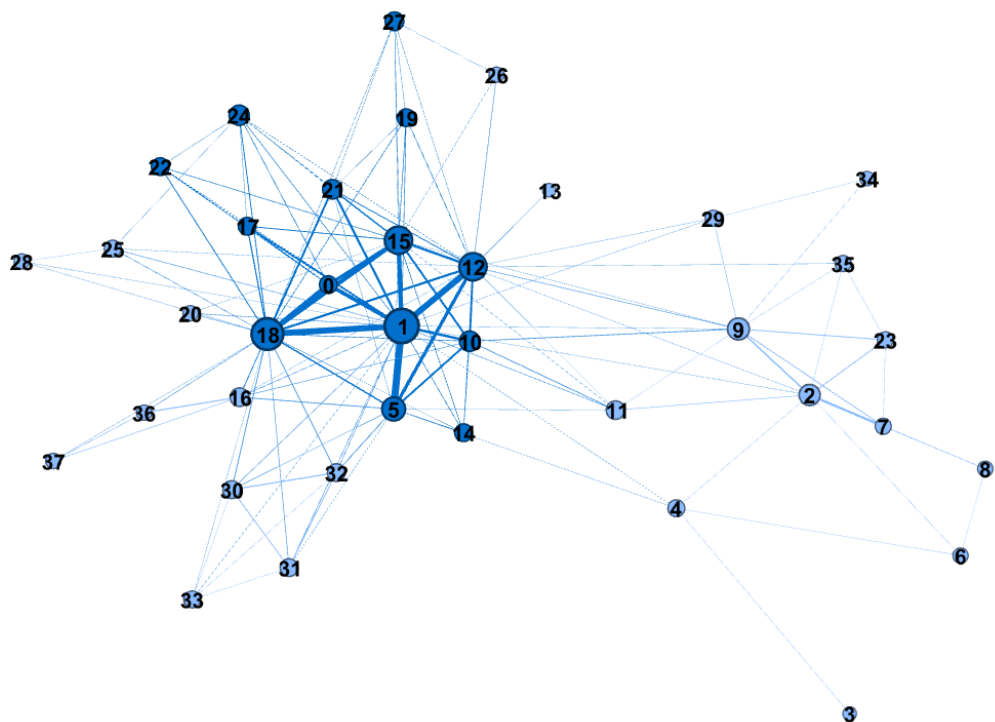
Слика 27



Слика 28

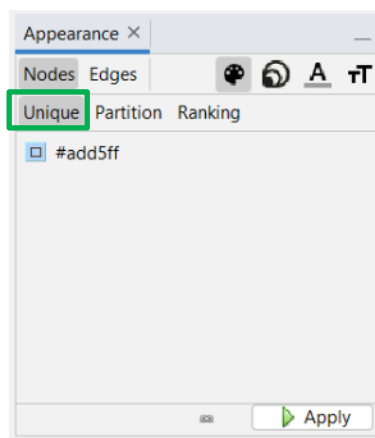
На слици 28 приказан је задати граф након клика на дугме *filter* за шел индекс 6 што је уједно и максималан шел индекс.

На слици 29 приказан је исти граф само након клика на дугме *select*, на њему су чворови у 6-соге обојени тамно плаво, док су остали обојени светло.



Слика 29

На сликама величина чвора зависи од степена чвора, чворови који су на слици већи имају и већи степен, како се подешава величина је већ описано у делу о метрици за просечну вредност степена мреже. Овакав приказ боја се постиже тако што пре примене филтера у одељку *appearance* изаберемо опцију *color* а затим бирамо опцију *unique* на слици 30 је уоквирена зеленом бојом, кликом на боју отвара се мени где можемо изабрати боју коју желимо. Да би се подешавања применила потребно је кликнути на *apply*.



Слика 30

Да би чворови са шел индексом били обојени тамније потребно је поновити исти поступак након примене филтера за $k\text{-core} = 6$.

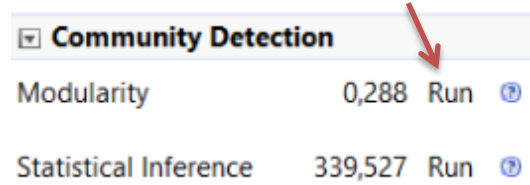
На слици 31 приказано је 14 чворова који имају шел индекс 6 у *data laboratory*.

Id	Label	Interval	name	Degree
0	0		R2-D2	6
1	1		QUI-GON	26
5	5		OBI-WAN	13
10	10		CAPTAIN P...	9
12	12		JAR JAR	18
14	14		BOSS NASS	6
15	15		PADME	18
17	17		WATTO	6
18	18		ANAKIN	23
19	19		SEBULBA	6
21	21		SHMI	8
22	22		C-3PO	6
24	24		KITSTER	9
27	27		JABBA	7

Слика 31

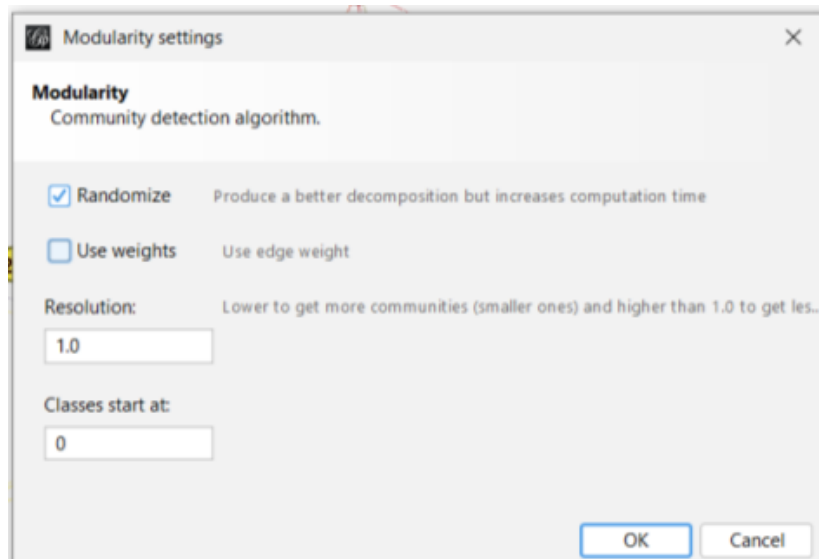
Community detection

Детекција заједница се врши тако што у одељку *statistics* кликне на *run* поред *modularity*, као што је приказано на слици 32. Након клика на *run* дугме отвара се нови прозор који је приказан на слици 33 у којем се могу подешавати неке опције.



Слика 32

Можемо подешавати од ког индекса креће нумерисање класа, резолуцију и слично. Не чекирамо опцију *use weights* јер не користимо тежине веза у графу, нису битне за анализу. Остављамо резолуцију на *default* вредности, као и почетни индекс за нумерисање класа. Када све поставимо како нам одговара кликнемо на ОК.



Слика 33

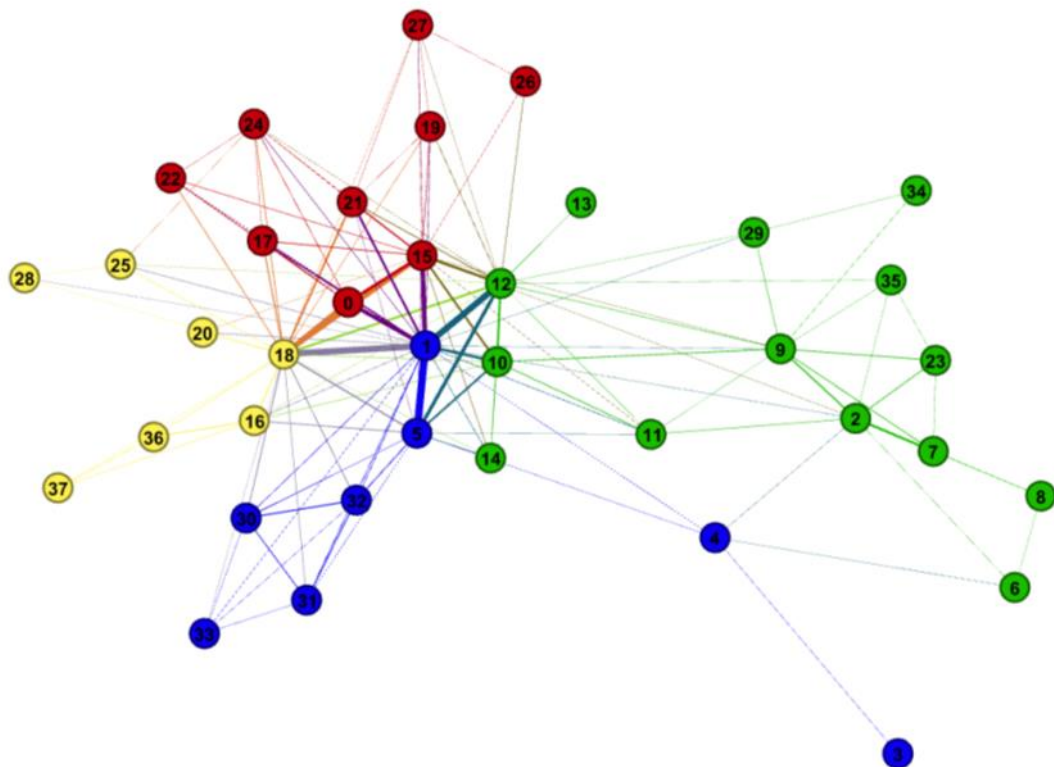
Након клика на дугме ОК добијамо извештај. На слици 34 приказане су заједнице на *x* оси и број чворова који припадају тој заједници на *y* оси.

Видимо да у овом задатку постоји 4 заједнице (0, 1, 2 и 3)



Слика 34

Са графикана се види да у заједници са индексом 0 има 14 чворова, у заједници са индексом 1 има 9 чворова, у заједници са индексом 2 има 7 чворова а у заједници са индексом 3 има 8 чворова.



Слика 35

На слици 35 приказан је граф где су чворови обојени тако да чворови исте боје припадају истој заједници. На слици 36 се може видети којом бојом су обојени чворови које заједнице.

Modularity Class		
■	0	(36,84%)
■	1	(23,68%)
■	3	(21,05%)
■	2	(18,42%)

Слика 36

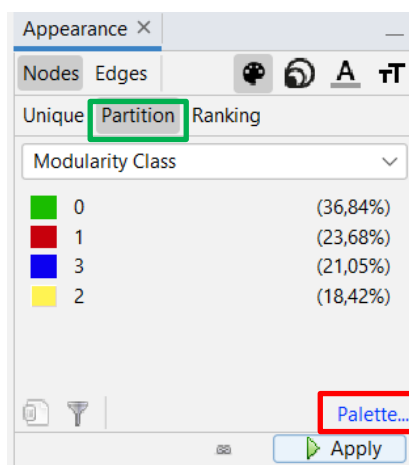
Заједница 0 (2, 6, 7, 8, 9, 10, 11, 12, 13, 14, 23, 29, 34, 35)

Заједница 1 (0, 15, 17, 19, 21, 22, 24, 26, 27)

Заједница 2 (16, 18, 20, 25, 28, 36, 37)

Заједница 3 (1, 3, 4, 5, 30, 31, 32, 33)

Приказ као на слици 35 постиже се тако што у одељку *appearance* након избора опција *nodes* и *color* бирамо опцију *partition* затим из падајућег менија бирамо опцију *modularity class*. Када то урадимо добићемо неке боје које ће одговарати заједницама у мрежи, кликом на *palette* можемо изабрати други понуђени спектар или дефинисати свој. Након одабира боја кликнути на *apply* као што је приказано на слици 37.



Слика 37